Reusing Software to Build Data Processing Systems: NPP Science Data Segment Case Study

Shahin Samadi, Ryan Gerard,
Mary Hunter, James J. Marshall
Innovim / NASA Goddard Space Flight Center
Mail Stop 614.5
Greenbelt, MD 20771
301-614-5457, 301-614-5458,
301-614-5518, 301-614-5452
shahin.samadi@gsfc.nasa.gov, ryan.gerard@gsfc.nasa.gov,
mary.hunter@gsfc.nasa.gov, james.marshall@gsfc.nasa.gov

Robert J. Schweiss
NASA Goddard Space Flight Center
Mail Stop 429
Greenbelt, MD 20771
301-286-1223
robert.j.schweiss@nasa.gov

Robert E. Wolfe and Edward J. Masuoka
NASA Goddard Space Flight Center
Mail Stop 614.5
Greenbelt, MD 20771
301-614-5508, 301-614-5515
robert.e.wolfe@nasa.gov, edward.j.masuoka@nasa.gov

Abstract—Over the years, numerous large and complex information systems have been created to store, process, and disseminate vast volumes of remotely-sensed Earth science data. These systems have the potential to be reused to process similar data from other missions or instruments, reducing risk, schedule, and associated development cost for future projects.[1][2]

One example of this kind of reuse is found in the Science Data Segment (SDS) of the National Polar-orbiting Operational Environmental Satellite System (NPOESS) Preparatory Project (NPP). This project provides remotely-sensed atmosphere, land, ocean, ozone, and sounder data that will serve the meteorological and global climate change to scientific communities while also providing risk reduction for NPOESS, the U.S. Government's future low-Earth orbiting weather and environmental satellite system. NPP serves as a bridge between current and future missions by providing pre-operational on-orbit test and risk reduction for key NPOESS instruments and ground-based data processing capabilities, while maintaining continuity of environmental data used for long-term climate change research. The role of the SDS is to independently assess the quality of NPP data products for accomplishing climate research.

The NPP Science Data Segment achieves its goals by leveraging off of existing processing centers. NPP SDS will perform the evaluation and analysis of atmosphere, land, ocean, ozone, and sounder data products. This paper will focus on the software reuse aspects of the SDS, and in particular, reuse of the Moderate Resolution Imaging Spectroradiometer (MODIS) Adaptive Processing System (MODAPS) in the development of the Land Product Evaluation and Algorithm Test Environment (PEATE) and Science Data Depository and Distribution Element (SD3E). Additional pieces of the SDS also reuse other existing Earth science data systems such as the Sea-viewing Wide Field-of-view Sensor (SeaWiFS) Ocean Data Processing System (ODPS) and the Ozone Monitoring Instrument (OMI) Data Processing System (OMIDAPS). Due to space constraints and to avoid duplication of content, this paper focuses on how the reuse of the existing MODAPS software system assists the NPP SDS project in meeting its requirements and goals.

TABLE OF CONTENTS

1

[1] 1-4244-0525-4/07/$20.00 ©2007 IEEE
[2] IEEEAC paper #1677, Version 3, Updated December 8, 2006

1. INTRODUCTION

Earth observing satellite missions have collected data about the Earth's climate and ecosystems for more than 30 years. In the beginning many of the data systems which produced global products for these missions were custom built with little effort at reuse. Over the last 5 years, with the advent of cost effective processing platforms (Linux systems running on commodity hardware, e.g. personal computers) the focus of the data processing teams and the science communities they support has shifted from developing single purpose systems to finding effective ways to further reduce cost and increase collaboration through developing multi-mission systems with a high degree of reuse to process data from related instruments in a single science discipline. Specifically, teams developing multi-mission systems have the following common goals:

- Ingest, archive, distribute, process, and analyze data products more efficiently

- Process large data sets with complex scientific algorithms in a seamless workflow

- Collaborate in a transparent environment by sharing data, processing algorithms, and pooling computer resources

- Spend less on technology infrastructure and software development and more on achieving operational and scientific objectives

The Earth science community faces many challenges in building a system that can meet the goals of most Earth observing missions. These challenges include:

- Data Quantity – Earth observing instruments generate large volumes of data daily. National Polar-orbiting Operational Environmental Satellite System (NPOESS) Preparatory Project (NPP) is projected to generate several terabytes of data on a daily basis.

- Distributed Knowledge Base – The Earth science community is dispersed both geographically and by discipline.

- Technology Infrastructure – Technology solutions supporting these missions are often complex and expensive.

- Software Development – Developing systems and applications supporting new missions often requires extensive domain-specific knowledge.

The NPP mission provides remotely-sensed atmosphere, land, ocean, ozone, and sounder data that service the meteorological and global climate change to scientific communities while also providing risk reduction for NPOESS, the U. S. Government's future low-Earth orbiting weather and environmental satellite system [1]. The NPP serves as a bridge between current and future missions by providing pre-operational on-orbit test and risk reduction for key NPOESS instruments and ground-based data processing capabilities, while maintaining continuity of environmental data used for long-term climate research [2].

One of the NPP mission segments is the Science Data Segment (SDS). The SDS principally provides the capabilities to assess and verify the quality of NPP products identified as Raw Data Records (RDRs), Sensor Data Records (SDRs), and Environmental Data Records (EDRs). The primary role of the SDS is to independently assess the quality of the NPP EDRs, also known as NASA Level 2 products, for accomplishing climate research.

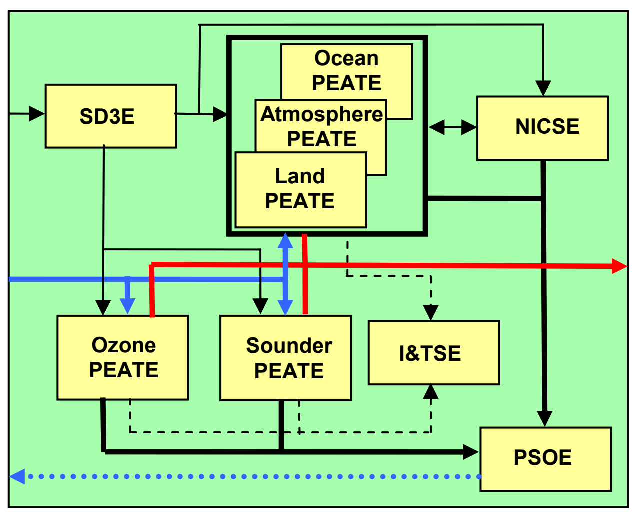

Figure 1 – Simplified SDS Block Diagram

As shown in Figure 1, the SDS is composed of nine elements: the SDS Data Distribution and Depository Element (SD3E), the Integration and Test System Element (I&TSE), the Project Science Office Element (PSOE), the NPP Instrument Calibration Support Element (NICSE), and five Product Evaluation and Analysis Tools Elements (PEATEs), one for each of the Atmosphere, Land, Ocean, Ozone, and Sounder disciplines. The NPP Science Data Segment achieves its goals by leveraging off of existing processing centers [3]. In particular:

- Land data analysis will use the Moderate Resolution Imaging Spectroradiometer (MODIS) Adaptive Processing System (MODAPS). MODAPS generates Level 2 through Level 4 MODIS Land science products, which are shipped to Earth Observing System Data and Information System (EOSDIS) Distributed Active Archive Centers (DAACs) for archival and to the MODIS science team for quality control [4].

- Ocean color and sea surface temperature product evaluation and analysis is accomplished by applying existing hardware and software from the Sea-viewing Wide Field-of-view Sensor (SeaWiFS) project which develops and operates a research data system that processes, calibrates, validates, archives, and distributes data received from Earth-orbiting ocean color sensors [5].

- Ozone products will be assessed by Ozone Monitoring Instrument (OMI) Data Processing System (OMIDAPS), a system that provides a framework for execution of science algorithms that transform lower level ozone data acquired from NASA's Aura satellite into higher level science data products [6].

- The University of Wisconsin Space Science and Engineering Center (SSEC) will evaluate atmosphere products. The Atmosphere PEATE is largely based on the Ocean Data Processing System (ODPS) [7].

- The Jet Propulsion Laboratory (JPL) will evaluate sounder products, and the Sounder PEATE will leverage the existing Atmospheric Infrared Sounder (AIRS) software and hardware [8].

- Additional pieces of the project such as the Science Data Depository and Distribution Element (SD3E) also rely on existing Earth science data systems. SD3E is primarily reusing components of MODAPS [9].

This paper provides an overview of the SDS software reuse in general and the Land PEATE and SD3E software reuse in particular. Due to space constraints and to avoid duplication of content, we are unable to fully cover the other systems, but their reuse is similar to the reuse of MODAPS described here. The lessons learned and success of the software reuse will be beneficial to driving down the cost and schedule of developing the NPP Science Data Segment. Since NPP will be one of the missions that will contribute to and participate in the Global Earth Observation System of Systems (GEOSS) collaboration, these lessons learned will also be valuable to the future missions that are included in the GEOSS plan.

2. Software Reuse and Reusability

Software reuse often is defined as utilizing existing software artifacts. Software reusability is not a new phenomenon; it has been demonstrated that utilizing existing software artifacts can significantly improve productivity and quality while decreasing the cost of software development. Large software systems are usually the results of the integration of many smaller components. Reusing software has many benefits such as increased productivity, reduced time to market, and improved quality. The motivation for reuse is based on productivity and quality improvements. Productivity can be defined as a function of cost and labor. If reuse can save cost and labor compared to developing software from scratch, then it enhances productivity. The reliability of reusable software has a direct impact on the quality of the system that is reusing them. In general, reusable software must be purposely designed and implemented. Reusable artifacts are often source code, but do not have to be. Table 1 lists a variety of software components that can be reused.

Table 1 – Potential Reusable Software Components

Reusable Software Components
Operational Source code
Analysis and design specifications
Plans (project management)
Data (testing)
Synthetic Data Generators, and Analysis Tools
Expertise / experience (life cycle model, quality assurance)
Information used to create software and documentation

It is important to complement the “what” information of reuse components with their “how” information in order to make any use of those components. Work products and services other than just source code are essential to facilitate reuse across platforms, products, and organizations. Documentation such as a specification or an installation guide is one example of an item that can be packaged with reusable components to subsequently improve their reusability.

If software components are designed with reusability as a requirement, they can be used repeatedly. However, even if the software was not designed with reuse in mind, if the software domain is similar enough across a broad range of application areas such as user interface, data structure, and sorting algorithms, then the software can be reused more easily. Software components can be reused in a similar application within the same problem domain; a domain analysis is required to achieve this reuse. Domain analysis is a process by which information used in developing software systems is identified, captured, and organized with the purpose of making it reusable when creating new systems. Domain analysis deals with the development and evolution of an information infrastructure to support reuse.

Software development process is important for the success of software reuse. A top-down and single project life cycle model such as the "waterfall" model might not be appropriate for software reuse. Software reuse is not inherently top-down and it needs a perspective which may be beyond the development of a single project. The reuse process needs to follow a structured approach [10] such as:

(1) Specifying the object to be created

(2) Searching the project, domain, and general databases for reuse candidates

(3) Evaluating the candidates to determine which (if any) should be used

(4) Estimating potential magnitude of effort

(5) Modifying, if necessary, to fit specific needs

(6) Integrating the reusable components

(7) Validating

(8) Feeding back the knowledge regarding the payoff of reuse

The SDS followed this approach in the developing of a distributed evaluation system for NPP from systems which process, archive and distribute Earth Observing System (EOS) data products. Once the SDS requirements, end-to-end mission concept, and logical system architecture were defined, it was possible to specify the systems to be created. Next, existing systems that could be reused were identified, and, from the overall set, five data systems were selected. These systems were then modified to meet mission-specific requirements. Mission-specific requirements included interfaces to external systems that provide or archive data products, product formats and metadata used by NPP, and adapting the software which produces each data product to run in the data system.

3. LAND PRODUCT EVALUATION AND ALGORITHM TEST ELEMENT (PEATE)

Background

The Land PEATE is responsible for the NPP Visible Infrared Imaging Radiometer Suite (VIIRS) SDR and EDR quality assessment, and demonstration of algorithm enhancements. The Land PEATE also supports analysis in evaluations performed by the NPP Instrument Calibration Support Element (NICSE) by providing access to data and tools for ad-hoc analysis of the VIIRS SDRs and algorithms. The Land PEATE will be hosted on an expanded MODAPS system. Some modifications and enhancements to the existing MODAPS processing framework are needed in order to support VIIRS product generation and evaluation. The bulk of the work involves adapting, integrating and testing production software for VIIRS Land products and adapting specific tools for the quality assurance of products.

The software being reused comes from the MODAPS system that was developed to produce global products for the Moderate Resolution Imaging Spectroradiometer (MODIS) instrument on the Terra and Aqua spacecrafts. Figure 2 is a block diagram of the existing MODAPS system indicating which components of MODAPS that were reused to create the different components of the Land PEATE and SD3E.

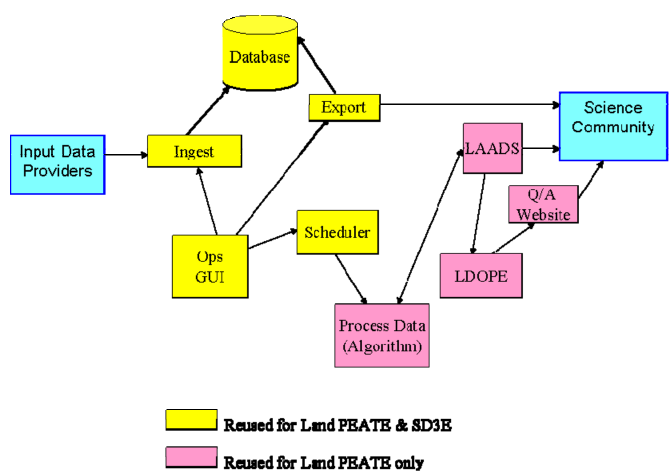

Figure 2 – MODAPS Block Diagram

Reusing MODAPS to produce VIIRS assessment products has proven simple since the VIIRS instrument and the associated algorithms has heritage from the MODIS instrument.

Prior to its use in NPP, the MODAPS system was reused on several projects to produce atmospherically corrected and Earth-located radiances for the Multispectral Scanner (MSS), Thematic Mapper (TM), and Enhanced Thematic Mapper Plus (ETM+) instruments onboard the various Land Remote-Sensing satellites (Landsats), and the Earth sensor Advanced Very High Resolution Radiometer (AVHRR) instruments on the NOAA 7 through NOAA 14 spacecraft. For each of these instruments, most of the development was spent in refining software to produce improved science products with only minor effort expended on modifying the overall processing and distribution framework of MODAPS.

Rationale

The reuse of the MODAPS system has greatly reduced the amount of new software development in the Land PEATE. If the software had to be developed from scratch, the time to develop the system would be more than tripled. The reuse has also proven to allow the scientists to have a more stable system. Many of the problems discovered with new development have already been resolved.

There are two main advantages to reusing MODAPS. The first is the fact that the system being built was similar enough to an existing system that reuse was possible. The VIIRS processing is different from the MODIS processing, but the VIIRS ingest, archive, distribution, and tools are very similar to MODIS, and therefore reuse was much easier than reusing a system built for an entirely different purpose. The second advantage of reusing the MODAPS is the resident knowledge available. These are explained in further detail in the Advantages of Reuse section below.

Land PEATE Reuse of MODAPS

MODAPS is being enhanced and augmented to support the NPP mission. The underlying core software will continue to be shared between missions. The core software components will be reused with approximately 97% of the software being used without modification. The next sections describe the MODAPS subsystems that are reused for NPP. This includes: Ingest, Operator (OPS) GUI, Database, Export, Scheduler, Land and Atmosphere Archive and Distribution System (LAADS), Land Data Operational Product Evaluation (LDOPE), Process Data, and Quality Assessment (Q/A) Website. The description of modifications for each subsystem and the reason for the modifications are included, as well as the estimated savings due to software reuse.

Ingest

The MODAPS system has software to ingest the data products as they arrive from the data distribution centers. This ingest is automated and triggered by the arrival of a data product. Ingest is reused, but modified to handle NPP-specific metadata and filename conventions of VIIRS RDRs, SDRs, EDRs, and ancillary data. The MODAPS ingest software currently ingests data from MODIS, AVHRR, and Landsat data processing centers. Table 2 depicts the software reuse for both ingest and metadata software subsystems in terms of Source Lines of Code (SLOC).

Table 2 – Ingest Software Reuse Analysis

Subsystem Name	Language	Reused SLOC	Total SLOC	Percent Reuse
Ingest	SQL, Perl, Shell, HTML	8399	9261	90.7%
Metadata	SQL, Perl, Shell, HTML	12114	13699	88.4%

The products for MODIS, AVHRR, and Landsat all have different file naming conventions, as will those from NPP VIIRS. MODAPS makes use of character strings within the file names to uniquely identify the specific granules in many places throughout its system. An example of a file name for the MODIS Level 2 Cloud Properties product from the EOS Terra mission is as follows: MOD06_L2.A2000031.1200.005.2006032103015.hdf. The file names minimally contain character strings that identify the type of product, the version of the product, the time when the data within the product granule were observed and acquired from the telemetry, and the time when the product granule was generated by the MODAPS. The specific format of these file names is shown in Figure 3.

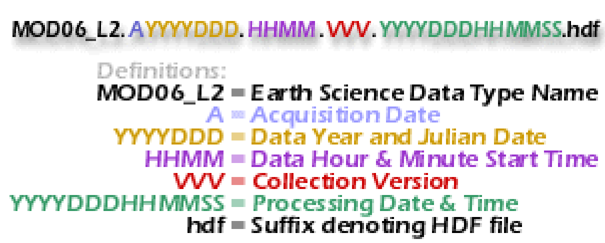

Figure 3 – Example of a Terra File Name Format

Operator (OPS) GUI, Database

The MODAPS database and operator (OPS) GUI subsystems are completely reused. The database software provides the framework for archiving and accessing the data products that are stored as a result of ingest. The database software is not coupled with the mission specific processing rules and therefore supports all missions with the same core software. The operator GUI is primarily there for manual intervention into the MODAPS system. Although the system is automated there are specific functions supported through the operator GUI, including routine database maintenance, and disk management. Table 3 shows the total source lines of code for the operator GUI and the database software subsystems, both with 100% reuse.

Table 3 – OPS GUI and Database Software Reuse Analysis

Subsystem Name	Language	Reused SLOC	Total SLOC	Percent Reuse
OPS GUI	Java, Perl, Shell, HTML	54994	54994	100%
Database	SQL, Perl, Shell, HTML	20207	20207	100%

Scheduler, Loader, and Data Production

The MODAPS scheduler software includes scripts that control the storage, processing, and distribution of the data products. Data storage involves loading meta-information to a database from file name and metadata files. Data processing is the sequence of software processing steps to generate the science products. Once all instrument data and ancillary data products are ingested into MODAPS, processing to higher level products begins. This processing step includes applying the science algorithms to the raw data records to produce the Sensor Data Records (SDRs), i.e., calibrated and Earth-located radiances, and applying algorithms to the SDRs to produce Environmental Data Records (EDRs), i.e., geophysical parameters such as land surface reflectance.

The MODAPS Scheduler, Loader, and Product Generation Executives (PGEs) are all required to produce SDR and EDR products for VIIRS. No modifications are needed for the Scheduler as it has already been used to process multiple missions (MODIS, AVHRR, and Landsat) without requiring any modifications. However, some changes to the Loader subsystem are required to capture differences between MODIS and VIIRS production rules. These production rules define which input products are staged for processing under different conditions and how to handle missing input products. While much of VIIRS processing is like MODIS product generation, feedback loops exist between gridded intermediate products in VIIRS and swath-based EDRs that are not found in MODIS swath-based (Level 2) products and, of course, there are minor differences in file naming conventions.

Product Generation Executives (PGEs) implement the science processing algorithms associated with each product from a given instrument. Of all elements of the MODAPS, PGEs are the ones with the least reuse between missions because they implement instrument-specific processing algorithms. Information to generate the Land VIIRS PGEs is obtained from the Operational production systems [11] and associated documentation. Additionally, the NPP Science Team members will provide alternative algorithms to assess the quality of the products over what is produced by the operational systems. The Land PEATE acquires the operational processing software from the production system and alternative software directly from Science Team members. The Land PEATE adapts operational algorithms to run as PGEs by replacing system calls that access memory objects in the NPOESS Interface Data Processing Segment (IDPS) with system calls to routines that read and write disk files in MODAPS. To promote reuse among PGEs for VIIRS processing, a common set of routines that read and write files in Hierarchical Data Format (HDF) – Earth Observing System (HDF-EOS) format using the HDF library will be developed and used to perform input/output operations for both the operational and Science Team developed PGEs. PGEs will then be integrated and baselined in MODAPS. Following this, the Land PEATE will endeavor to replicate a subset of the production SDRs and EDRs for subsequent assessment that will be compared with other instrument observations such as MODIS, or independently generated products. Table 4 summarizes the reuse of the Scheduler, Loader, and Product Generation subsystems.

Table 4 – Scheduler, Loader, Product Generation Software Reuse Analysis

Subsystem Name	Language	Reused SLOC	Total SLOC	Percent Reuse
Scheduler	SQL, Perl, Shell, HTML	6771	6771	100%
Loader	SQL, Perl, Shell, HTML	8945	9011	99.3%
Product Generation	Shell, Perl, HTML	4631	4317	100%

Land Data Operational Product Evaluation (LDOPE)

The Land Data Operational Product Evaluation (LDOPE) subsystem will be reused with changes to assess NPP SDRs and EDRs. Specifically, the Science Community Interface website will be modified to support the new NPP products, and quality assurance tools will be updated to support quality assurance of VIIRS SDR and EDR products. Table 5 is a list of four tools; a complete list can be found at the LDOPE web site [12].

Table 5 – LDOPE QA Tools

Tool Name	Description
convert_l1b_data	Convert MODIS L1B data to Top of Atmosphere (TOA) reflectance for the MODIS reflective bands and TOA radiance for the MODIS emissive bands and write these to 2D HDF SDS(s) that can be read by commercial off-the-shelf (COTS) software. The conversion is performed using the scale and offsets defined in the MODIS L1B product metadata
geolocation	Compute the geographic latitude and longitude of a MODIS Land L2G/L3/L4 pixel coordinate.
tile_id	Compute the MODIS Land L2G/L3/L4 tile id for a given latitude and longitude. This tool identifies the MODIS Land tile that corresponds to a known geographic location.
mosaic_sds	Create a spatial mosaic from different L3 MODIS Land products.

The quality assurance tools, processes, and applications currently being used in the quality assessment (Q/A) of MODIS Land products will be adapted to the VIIRS Land product formats and to meet the specific needs of the VIIRS Science Team. Specifically, the create browse images application will be modified to accommodate the new HDF formats and VIIRS Image bands. The time series plots used to perform outlier identification will be modified from existing MODIS source code. The current metadata database used for the evaluation of existing MODIS and AVHRR products will be adapted to identify the required fields in the context of NPP VIIRS Land SDR and EDR formats. Table 6 shows the software reuse for the LDOPE web site.

Table 6 – LDOPE Tools Software Reuse Analysis

Subsystem Name	Language	Reused SLOC	Total SLOC	Percent Reuse
QA Website	SQL, Mason, Java, Perl, HTML	84894	93306	90.9%

Land and Atmosphere Archive and Distribution System (LAADS)

The Land PEATE shares VIIRS products produced in MODAPS through the subsystem Land and Atmospheres Archive and Distribution System (LAADS). The LAADS provides a search and order interface, distribution mechanisms (FTP and Open-source Project for a Network Data Access Protocol (OPeNDAP)), virtual products via production on-demand, and custom processing of existing products for delivery to the VIIRS Land Science Team [13]. In addition to allowing users (restricted to NPP Science team for NPP) to search and order data from the archive, the web system is augmented to provide reports on NPP data production. The archive subsystem maintains the stored data products in the database and minor modifications were required to support the VIIRS data product storage. The export subsystem was reused to deliver the data products available to the science community. Table 7 summarizes the software reuse for LAADS, Archiver, and Export software subsystems.

Table 7 – LAADS Software Reuse Analysis

Subsystem Name	Language	Reused SLOC	Total SLOC	Percent Reuse
LAADS	Mason, HTML	648	648	100%
Archiver	SQL, Shell, Perl	13010	13644	95.4%
Export	SQL, Shell, Perl, HTML	21437	18093	100%

Advantages of Reuse

Using the Constructive Cost Model (COCOMO) [14], the estimated level of effort to develop the software without reuse is much greater. Table 8 shows that the software development effort would have been ten times the effort without software reuse. The effort with reuse is an estimate that does not include integration and testing, but the savings in development alone is noteworthy.

Table 8 – Estimated Effort with Software Reuse

Subsystem	With Reuse [staff-months]	Without Reuse [staff-months]
Ingest	3	25

Metadata	4	37
OPS GUI	0	161
Database	0	56
Scheduler	0	18
Loader	1	24
Product Generation	12	30
LDOPE	24	281
LAADS	0	1
Archiver	2	37
Export	8	50
Total	54	720

Another advantage of reusing the MODAPS is the resident knowledge available. The developers and scientists that currently support MODAPS are the same developers and scientists supporting the NPP VIIRS assessment. This Land PEATE team has a history of collaboration and developers and scientists meet frequently and work closely together. They are rich in their knowledge of the Earth science domain and have many experiences and expertise to leverage.

4. SCIENCE DATA DEPOSITORY AND DISTRIBUTION ELEMENT (SD3E)

Background

In the NPP SDS domain, the SD3E is responsible for ingesting, staging, and distributing the NPP data products. NPP SDS will obtain data products from NOAA's Comprehensive Large Array-data Stewardship System (CLASS), NPOESS's Interface Data Processing Segment (IDPS), and the NPOESS Science Investigator-led Processing System (NSIPS). All three data providers will be located at NOAA in Suitland, Maryland. The data products, once ingested from these providers, will be staged for distribution to the PEATEs and NICSE. SD3E is a new implementation mostly reusing MODAPS components.

The SD3E is composed of an FTP server, a database, and four software subsystems: the Scheduler, Ingest Controller, Interface Controller, and Utilities. The Scheduler manages all tasks and processes. The Ingest Controller verifies received data products (via checksum or digital signature) and makes those products available to the PEATEs via FTP server. The Interface Controller handles ad-hoc or subscription requests by the PEATEs for products residing on IDPS or CLASS. The Utilities subsystem includes a suite of tools that monitor the health of and maintain the SD3E.

Rationale

The set of functions, interfaces, and data formats used for the SD3E is different from that of previous systems. However, the general similarity of this system to the existing MODAPS system allowed for a high level of reuse. The scheduler, database, and export functions are essentially the same in the SD3E as in MODAPS, and approximately 53% of the code in the SD3E was reused from MODAPS with very little or no modification. This reduces the amount of time required for development.

SD3E Reuse of MODAPS

The SD3E makes use of the scheduler, export, database, and operator GUI components of MODAPS, as shown in Figure 4. Modifications to the components are necessary to meet varying requirements on how the data store is managed, including the duration of product availability and when data products expire and are removed from the data store, and to handle the HDF format data files that will be used by NPP.

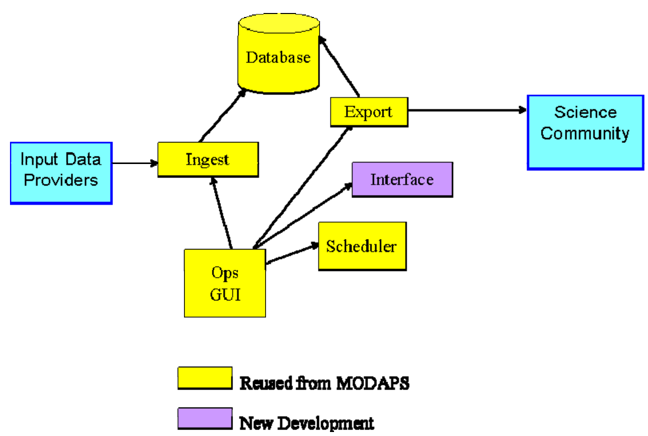

Figure 4 – SD3E Subsystem Reuse

SD3E reuses the architecture established by MODAPS, namely having a local inbound FTP directory for product and data delivery report retrieval and a corresponding local output FTP directory for PEATE access. Additionally, the SD3E hardware architecture decisions (disk controller type, Redundant Array of Independent Disks (RAID) configuration, server vendor) were made based on experiences of the MODPAS team.

Ingest, Operator GUI, Scheduler, Export, & Database

The MODAPS subsystems that were reused in the SD3E are: Ingest, Operator (OPS) GUI, Scheduler, Export, and Database data ingest subsystem was reused to create the

SD3E data ingest system. The ingest subsystem is comprised of a scheduler, database, ingest controller, and interface controller. All of these components were reused in the SD3E design with the modifications required to interface with the new external interfaces and meet mission-specific requirements. These include ingesting NPP data products, supporting mission-specific verification methods such as integrity files, parsing mission-specific metadata and storing it in the database, and locally storing NPP data products for 32 days. However, the overall workflow has not been altered from the MODAPS implementation.

Reuse Analysis

Much of the development of SD3E software is written in the Perl programming language, and Table 9 summarizes the Perl software source lines of code reused.

Table 9 – Perl Software Reuse Analysis

Total SLOC	4944
Total Reused SLOC	2611
Total Percent Reused	52.8%

While 52.8% of the total Perl software was reused, some subsystems were written from scratch, diluting the actual amount of reuse within SD3E. However, the Scheduler subsystem and Utilities subsystem include very high levels of reuse, as shown in Table 10.

Table 10 – Significantly Reused Perl Subsystems

Subsystem Name	Language	Reuse d SLOC	Total SLOC	Percen t Reuse
Scheduler	Perl	1190	1322	90%
Utilities	Perl	1421	1886	75%

To fit the SDS design, the Scheduler subsystem was tailored to use a PostgreSQL database rather than a Sybase database. The Scheduler will execute any task from any module as long as the task is defined. Necessary fields required to define a task include location of the executable, desired options, and frequency to be executed.

The Utilities subsystem monitors the system and notifies the system administrator should problems arise. For this subsystem, the disk consistency, disk management, and log management routines were reused with little change. New routines and functionality accounted for the bulk of the new code, such as reading metadata from HDF files.

The SQL software was analyzed separately from the Perl software. Table 11 illustrates the software reuse in SQL.

Table 11 – SQL Software Reuse Analysis

Total SLOC	1299
Total Reused SLOC	697
Total Percent Reused	53.6%

The SQL scripts are core scripts that control the process flow of the SD3E processing. These scripts were not written with the intent of being reused, but were written with standard SQL. This allowed some of them to be completely reused without modification. Table 12 highlights areas of significant reuse relating to the SD3E database.

Table 12 – Significantly Reused Database Subsystems

Subsystem Name	Language	Reused SLOC	Total SLOC	Percent Reuse
Table Definitions	SQL	288	432	66.6%
Stored Procedures	SQL	409	625	65%

Approximately 61% of the database tables are reused from MODAPS. A few new tables are required to support the NPP mission. These tables support the new external interfaces (IDPS and CLASS).

The stored procedures control the database activities for inserting products into the database and storing the various states of the data product requests. SD3E inherited a subset of the database schema from MODAPS. In particular, SD3E tracks its disk and file metadata using the same states. The SD3E is able to benefit from MODAPS design, which functioned for many years without failure.

One of the lessons learned from MODAPS was that traditional systems built with commercial off-the-shelf (COTS) database applications were costly and required reoccurring annual vendor maintenance contracts, which were a burden to sustain and maintain. Therefore, an open source database was chosen for use in the SD3E.

It is also important to note that not all software reuse can be shown through a static analysis. In fact, subsystems like the Ingest Controller and Interface Controller are written from scratch, but call many functions and routines that were reused from other subsystems. A dynamic analysis would show that the bulk of runtime functions executed are reused, while a significant portion non-reused code is executed for anomalous cases.

Advantages of Reuse

Using the Constructive Cost Model (COCOMO) [14], the effort required was estimated from the project size (using SLOC from Tables 9 and 11). The effort without reuse is noticeably greater. As shown in Table 13, the software development effort without reuse is just over double the effort with reuse.

Table 13 – Estimated Efforts for SD3E

Language	With Reuse [staff-months]	Without Reuse [staff-months]
Perl	6	13
SQL	1.5	3
Total	7.5	16

By reusing existing code, less new code has to be written. Also, existing code has been tested previously, thus reducing the amount of new testing that must be done, while helping to ensure the reliability of the new system. In addition to reusing code, reusing the lessons learned from the MODAPS project and the experience of its developers provides benefits by helping the SD3E make better informed choices.

5. CONCLUSIONS

The NPP SDS is a good example of software reuse. The SDS has all of the challenges associated with developing large and complex systems in support of Earth observing instruments. The reuse of existing systems with some new development and some modification greatly reduces the level of effort, schedule, and risk to create a system for NPP SDS. The resources that accompany the reused systems, such as people, hardware, knowledge, testing, data, and software contribute to the reduction in effort and risk. The estimated effort to build the system from scratch would be more than double the current effort, which includes reuse. The fact that the developers modifying the software being reused have the knowledge of previous missions allows for the increased production rate and decreased effort and duration. The developers are able to get a jump start on the modifications by testing the modified software with existing mission data rather than having to wait for simulated or actual NPP mission data to exist in the right format. Although the modifications include software that will handle the new format, the scientists are able to verify portions of their software using existing data formats.

The ability to collaborate among the reused systems also facilitates the reuse process. Experienced developers can explain non-intuitive features to the newer developers. Many duplicate efforts can be costly and time consuming. The collaboration of all of the developers that are modifying their existing systems greatly helps the development effort. Sharing the knowledge of one developer with others can save developer time. The MODAPS, SeaWiFS, and OMIDAPS developers all working together with a common data format and a common data interface (SD3E) help facilitate the reuse process.

NPP and NPOESS represent the U.S. Government's next generation, polar-orbiting, Low Earth-Orbit (LEO) satellite constellation and end-to-end systems for environmental remote sensing. NPOESS represents a major portion of the LEO component of a "Future National Operational Environmental Satellite System" which could be envisioned as a significant contributing component moving towards a Global Earth Observation System of Systems (GEOSS) [15]. Even though the reuse of MODAPS for NPP is made easy by the similarity of the source and target systems, the general functions of NPP SDS software will be applicable to systems that are included in GEOSS. Many of the same obstacles such as data processing, data distribution, and data storage, will exist at even a greater magnitude in the future. The lessons learned from this example of reusing existing software will be valuable to the future missions that are included in the GEOSS plan. Regardless of data formats and the processing that the data may require, the reuse of software design and application can reduce the effort and cost of future missions. The GEOSS vision of distributing data worldwide and storing data indefinitely will greatly benefit from the NPP SDS expertise.

The additional software effort required to accommodate different data formats and how to handle the various standards in data formats provide lessons that will be shared with the developers and scientists that support GEOSS missions. Evolved data standards combined with lessons learned from adapting comparison software will also be utilized during the GEOSS era. The contributing systems will range across the processing cycle, from primary observation to information production. Through GEOSS, they will share observations and products with the system as a whole, and will take the necessary steps to ensure that the shared observations and products are accessible, comparable, and understandable, by supporting common standards and adaptation to users' needs [16]. The missions that are included in GEOSS will have additional requirements to share data among different nations, with particular urgency for near real-time data distribution, so realizing the benefits of existing expertise will allow future developers to devote the time and effort necessary to accommodating these additional requirements.

ACKNOWLEDGMENTS

The authors thank the Land PEATE members, including Gary Fu and Gang Ye, for their help in providing the software information and for their assistance in helping to understand the MODAPS system.

REFERENCES

[1] Wilczynski, P.A., "The National Polar-orbiting Operational Environmental Satellite System (NPOESS) Preparatory Project (NPP): leading the change for NPOESS," *Geoscience and Remote Sensing Symposium, 2003. IGARSS '03. Proceedings. 2003 IEEE International*, vol.1, no.pp. 360-362 vol.1, 21-25 July 2003

[2] NPP web site http://jointmission.gsfc.nasa.gov/

[3] Schweiss, R.J.; Ho, E.; Ullman, R.; Samadi, S., "The NPOESS Preparatory Project Science Data Segment: Brief Overview", *Geoscience and Remote Sensing Symposium, 2006. IGARSS '06. Proceedings. 2006 IEEE International*, in preparation

[4] Masuoka, E.; Tilmes, C.; Devine, N.; Gang Ye; Tilmes, M., "Evolution of the MODIS science data processing system," *Geoscience and Remote Sensing Symposium, 2001. IGARSS '01. IEEE 2001 International*, vol.3, no.pp. 454-1457 vol.3, 2001

[5] SeaWiFS Project Home Page web site http://oceancolor.gsfc.nasa.gov/SeaWiFS/

[6] Tilmes, C.; Linda, M.; Fleig, A.J., "Development of two Science Investigator-led Processing Systems (SIPS) for NASA's Earth Observation System (EOS)," *Geoscience and Remote Sensing Symposium, 2004. IGARSS '04. Proceedings. 2004 IEEE International*, vol.3, no.pp. 2190-2195 vol.3, 20-24 Sept. 2004

[7] Space Science and Engineering Center, University of Wisconsin-Madison web site http://www.ssec.wisc.edu/

[8] Atmospheric Infrared Sounder (AIRS) web site http://www-airs.jpl.nasa.gov/

[9] Schweiss, R.J., Samadi, S., Ho, E.L., Gerard, R. (2005), An Overview of the NPOESS Preparatory Project (NPP) Science Data Segment, *Eos Trans. AGU, 86*(52), Fall Meet. Suppl., Abstract IN23B-1210

[10] Earth Science Data System Software Reuse Working Group web site http://softwarereuse.nasa.gov/

[11] Kevin Souza, Chad Fox, Kerry Grant, and Scott Turek, NPOESS Interface Data Processing Segment (IDPS) architecture and software, Proc. SPIE Int. Soc. Opt. Eng. 5549, 28 (2004)

[12] Land Quality Assurance web site http://landweb.nascom.nasa.gov/QA_WWW

[13] MODAPS web site https://modaps.nascom.nasa.gov/

[14] Boehm, Barry W., Software Engineering Economics, Prentice Hall, 1981.

[15] Mango, Stephen A., Introductory Remarks: Future National Operational Environmental Satellites for the American Meteorological Society Annual Technology Program conference in 2005, http://ams.confex.com/ams/Annual2005/techprogram/paper_89524.htm

[16] The Global Earth Observation System of Systems (GEOSS) 10-Year Implementation Plan *(As adopted 16 February 2005)*

BIOGRAPHY

Shahin Samadi *is a senior staff member of Innovim supporting NASA Goddard Space Flight Center and an adjunct faculty member of Johns Hopkins University. He has over 20 years of experience in architecting large-scale data processing systems. He has presented and published many technical papers in the area of software and systems engineering. Mr. Samadi has a Ph.D. and M.S. in Computer Science and a B.S. in Electrical Engineering from the George Washington University.*

Robert Schweiss *is currently supporting the NASA NPP Project Office in the capacity as Science Data Segment Manager at NASA Goddard Space Flight Center. He has over 18 years of professional experience with 15 years as a NASA Civil Servant. He has held key system engineering and leadership positions on the development of the Ground Data Handling Systems for the Landsat 7 and Triana Missions. Most recently he was the Calibration Validation manager for the Landsat Data Continuity Mission. Mr. Schweiss has a B.S. in Computer Science from the University of Maryland and a M.S. in Computer Science from the George Washington University.*

Edward Masuoka *is the Head of the Terrestrial Information Systems Branch at NASA Goddard Space Flight Center. For the last 10 years, he has led a team of civil servants and contractors developing software and science processing systems for the MODIS (Moderate-resolution Imaging Spectroradiometer) instruments on the Earth Observing System's Terra and Aqua spacecraft. Current projects his team is pursuing include improving data access and distribution of MODIS products and the generation of well-calibrated long-term data sets from Earth-observing satellite instruments.*

Robert Wolfe *works in the Terrestrial Information Systems Branch at NASA Goddard Space Flight Center. For over 10 years he has led MODIS geolocation activities, coordinated the MODIS Land science discipline activities and participated in the evolution of the MODAPS system. He is currently co-chair of NASA's Earth Science Data System Software Reuse Working Group, is NASA's Terra Deputy Project Scientist for Data, is a key participant in the Land PEATE and NICSE activities, and participates in project studying terrestrial biophysics over North America. Mr. Wolfe has B.S. degrees in physics and mathematics from Bridgewater College and over 25 years of experience in Earth remote sensing.*